Statistical Process Control

MSc: Statistics and Actuarial-Financial Mathematics

Konstantinos Bourazas
kbourazas@aegean.gr

Department of Statistics and Actuarial-Financial Mathematics

Course #6

March 26, 2026

Process capability analysis

- A process can be **IC** and still produce parts that fail to meet customer requirements
- For a quality characteristic X, the customer specifies a **Lower Specification Limit (LSL)** and an **Upper Specification Limit (USL)**
- A product is **conforming** if $\text{LSL} \le X \le \text{USL}$ and **non-conforming** otherwise
- The most direct measure of capability is the conforming probability

$$P(\text{LSL} \le X \le \text{USL}) \overset{X \sim N}{=} P\left(\frac{\text{LSL} - \mu}{\sigma} \le Z \le \frac{\text{USL} - \mu}{\sigma}\right)$$

- This probability is easy to interpret but does not separate the roles of **spread** and **centering**, which is why we also use **process capability ratios (PCRs)**

The C_p ratio

- C_p compares the specification width to the natural process spread

$$C_p = \frac{\text{USL} - \text{LSL}}{6\sigma}$$

- The denominator 6σ is the width of the interval $(\mu - 3\sigma,\ \mu + 3\sigma)$, which contains 99.73% of a normal distribution
- Interpretation
  - $C_p > 1$ means the specifications are wider than the process spread (desirable)
  - $C_p = 1$ means they are exactly equal
  - $C_p < 1$ means the process spread exceeds the specification width
- **Limitation**: C_p depends only on σ and ignores the process mean μ. Two processes with the same spread but different centering give the same C_p

The C_{pk} ratio

- To account for centering we define two one-sided ratios and take the smaller

$$C_{pl} = \frac{\mu - \text{LSL}}{3\sigma}, \qquad C_{pu} = \frac{\text{USL} - \mu}{3\sigma}, \qquad C_{pk} = \min(C_{pl},\ C_{pu})$$

- C_{pk} reflects **both** the spread and how far the mean is from the nearest specification limit
- Relationship with C_p

$$C_{pk} = \left[1 - \frac{|T - \mu|}{d}\right] C_p, \qquad T = \frac{\text{LSL} + \text{USL}}{2}, \quad d = \frac{\text{USL} - \text{LSL}}{2}$$

- Two consequences: $C_{pk} \leq C_p$ always, with equality only when $\mu = T$, and C_{pk} decreases as $|\mu - T|$ grows

The C_{pm} and C_{pkm} ratios

- An alternative way to penalise departure from target

$$C_{pm} = \frac{\text{USL} - \text{LSL}}{6\,\tau}, \qquad \tau = \sqrt{\sigma^2 + (\mu - T)^2} = \sqrt{E[(X - T)^2]}$$

- Since τ combines spread and off-centering, $C_{pm} \leq C_p$ with equality only when $\mu = T$. Equivalently

$$C_{pm} = \frac{C_p}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}$$

- For even greater sensitivity we replace C_p by C_{pk}

$$C_{pkm} = \frac{C_{pk}}{\sqrt{1 + \left(\frac{\mu - T}{\sigma}\right)^2}}$$

- C_{pkm} penalises off-centering more strongly than either C_{pk} or C_{pm} alone

Example 3.8 Comparing the PCRs

Assume LSL = -3, USL = 3, so $T = 0$ and $d = 3$

Same centre, different spread: $N(0, 1)$ vs $N(0,\ 0.5^2)$

	$N(0,1)$	$N(0,0.5^2)$
$C_p = C_{pk}$	1.000	2.000

When $\mu = T$ all PCRs coincide. Less spread $\Rightarrow$ higher capability

Same spread, different centre: $N(0, 1)$ vs $N(1, 1)$

	$N(0,1)$	$N(1,1)$
C_p	1.000	1.000
C_{pk}	1.000	0.667
C_{pm}	1.000	0.707
C_{pkm}	1.000	0.471

C_p cannot detect the off-centering. C_{pkm} is the most sensitive to it

Example 3.7 Estimating PCRs from data

- In practice μ and σ are replaced by $\bar{X}$ and s from an IC dataset
- Injection molding, $n = 50$ parts, $\bar{X} = 80.194$, $s = 2.775$, LSL $= 75$, USL $= 85$, $T = 80$

$$\hat{C}_p = \frac{10}{6 \times 2.775} = 0.601, \qquad \hat{C}_{pl} = 0.624, \qquad \hat{C}_{pu} = 0.577$$

$$\hat{C}_{pk} = 0.577, \qquad \hat{C}_{pm} = 0.600, \qquad \hat{C}_{pkm} = 0.576$$

- All estimates are well below 1, confirming **poor capability** even though the process is IC
- The mean is close to target ($\bar{X} \approx T$), so the problem is **excessive variability** relative to the specification width

Type I and Type II errors in control charts

- At each time point the chart performs an implicit hypothesis test

$$H_0: \text{process is IC} \qquad \text{vs} \qquad H_1: \text{process is OOC}$$

- **Type I error (false alarm)**: the chart signals when the process is actually IC

$$\alpha = P(\text{signal} \mid \text{IC})$$

- **Type II error (missed detection)**: the chart does not signal when the process is actually OOC

$$\beta = P(\text{no signal} \mid \text{OOC})$$

- In control charts we do not fix α at 0.05 as in classical testing. The conventional choice is $\alpha = 0.0027$ (from the 3σ limits), which keeps false alarms rare because the test is repeated at every single time point
- The trade-off is familiar: decreasing α makes false alarms rarer but increases β, so real shifts take longer to detect

The IC Average Run Length (ARL_0)

- Let N be the **run length**, the number of subgroups until the chart gives its first signal
- When the process is IC, each subgroup triggers a false alarm independently with probability α, so
$$N \sim \text{Geometric}(\alpha), \qquad P(N = k) = (1 - \alpha)^{k-1}\,\alpha, \quad k = 1, 2, \ldots$$
- The mean and standard deviation of this distribution are
$$\text{ARL}_0 = E[N] = \frac{1}{\alpha}, \qquad \sigma_{\text{RL}} = \frac{\sqrt{1-\alpha}}{\alpha} \approx \frac{1}{\alpha}$$
- For the conventional 3σ limits, $\alpha = 0.0027$, so $\text{ARL}_0 \approx 370$. The chart signals on average once every **370 subgroups** when IC
- A larger ARL_0 means fewer false alarms. We want ARL_0 to be **large** under IC and **small** when the process shifts

Multiple testing in Phase I and the FWER

- In Phase II the chart monitors subgroups one at a time, so the per-subgroup false alarm rate α is the relevant quantity
- In Phase I we evaluate n subgroups **simultaneously**, so even if the process is IC throughout, the probability of at least one false alarm is

$$\tilde{\alpha} = P(\text{at least one false alarm in } n \text{ subgroups}) = 1 - (1 - \alpha)^n$$

- This is exactly the **Family-Wise Error Rate (FWER)** from multiple testing
- For $\alpha = 0.0027$ and typical Phase I sizes

$$n = 10 \;\Rightarrow\; \tilde{\alpha} \approx 2.7\%, \qquad n = 20 \;\Rightarrow\; \tilde{\alpha} \approx 5.2\%, \qquad n = 50 \;\Rightarrow\; \tilde{\alpha} \approx 12.6\%$$

- So even with conservative 3σ limits, spurious signals in Phase I are not negligible when n is large
- If we want to control the overall $\tilde{\alpha}$, we can set the per-subgroup rate as $\alpha = 1 - (1 - \tilde{\alpha})^{1/n}$

The OOC Average Run Length (ARL_1)

- Under a sustained shift of size δ the probability that a single subgroup triggers a signal is $1-\beta(\delta)$, so the OOC run length again follows a geometric distribution

$$\text{ARL}_1(\delta) = \frac{1}{1-\beta(\delta)}$$

- This is the **zero-state** ARL_1, because it assumes the shift is present from the very first observation
- In practice the shift may occur at some unknown time $\nu > 0$ after monitoring has started. The **steady-state** ARL_1 averages over all possible change points and accounts for the information accumulated before the shift
- For a memoryless chart like the Shewhart chart, the zero-state and steady-state ARL_1 are the same, because the chart statistic at time i depends only on the data at time i
- For charts with memory (CUSUM, EWMA), the two can differ substantially

Conditional Expected Delay and the minimax criterion

- Suppose the shift occurs at time ν. The **detection delay** is $N - \nu$, where N is the time of the first signal after the shift
- The **Conditional Expected Delay (CED)** is

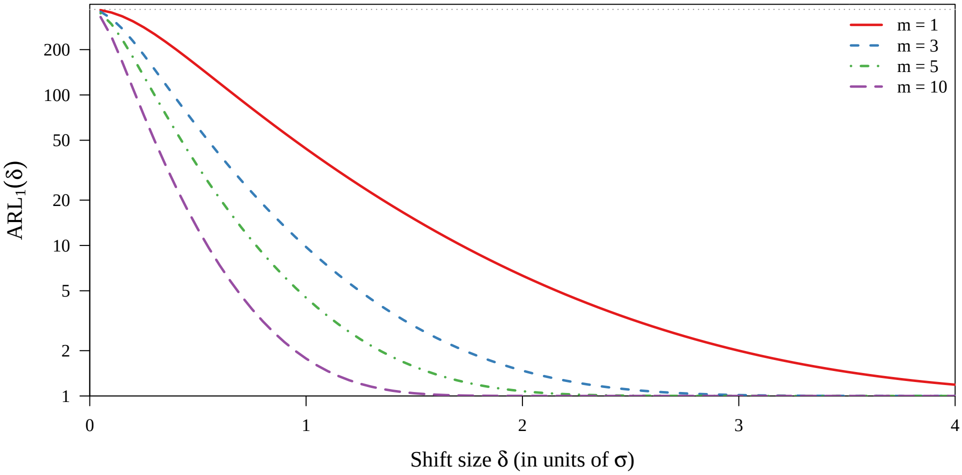
$$\mathrm{CED}(\nu) = E\left[N - \nu \,\middle|\, N \geq \nu\right]$$

  It measures how quickly the chart reacts, given that it has not yet signalled by the time the shift appears
- The CED can depend on ν for charts with memory, because the chart statistic at time ν depends on the path up to that point
- A natural optimality criterion is **minimax**: find the chart that minimises the **worst-case CED** over all possible change points

$$\inf_{\text{charts}} \sup_{\nu \geq 1} \mathrm{CED}(\nu) \qquad \text{subject to} \quad \mathrm{ARL}_0 \geq \gamma$$

- For the Shewhart chart the CED is constant in ν and equals the zero-state ARL_1, so the distinction does not arise

ARL_1 as a function of the shift size



Alternative performance metrics

- The ARL is the most widely used metric, but it has limitations. Because the geometric distribution is highly skewed, the mean may not be the best summary
- **Median Run Length** (MRL)
  - MRL_0 and MRL_1 are the medians of the IC and OOC run-length distributions
  - More robust than the ARL to the skewness of the geometric distribution and easier to interpret as "the time by which the chart has a 50% chance of signalling"
- **Average ARL** (AARL_1)
  - Instead of evaluating ARL_1 at a single shift δ, the AARL averages it over a distribution of plausible shift sizes
  - Useful when the exact magnitude of a potential shift is not known in advance
- In practice, ARL_0 and ARL_1 remain the standard benchmarks, but reporting the MRL alongside them gives a more complete picture